NASA's EOSDIS Cumulus

∷

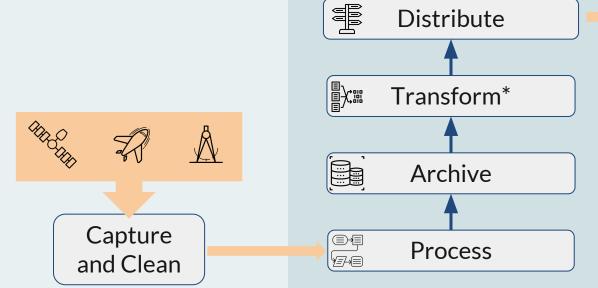**Ingesting, Archiving, Managing, and Distributing**

Earth Science Data from the Commercial Cloud

Katie Baynes - Civil Servant, NASA Goddard
Rahul Ramachandran - Civil Servant, NASA Marshall
Dan Pilone, Patrick Quinn, Jason Gilman - Element 84, Inc
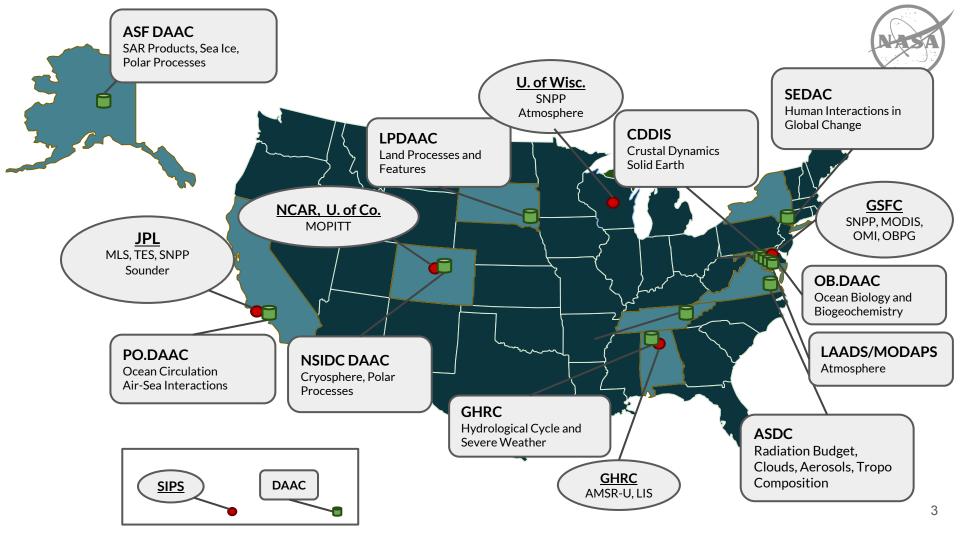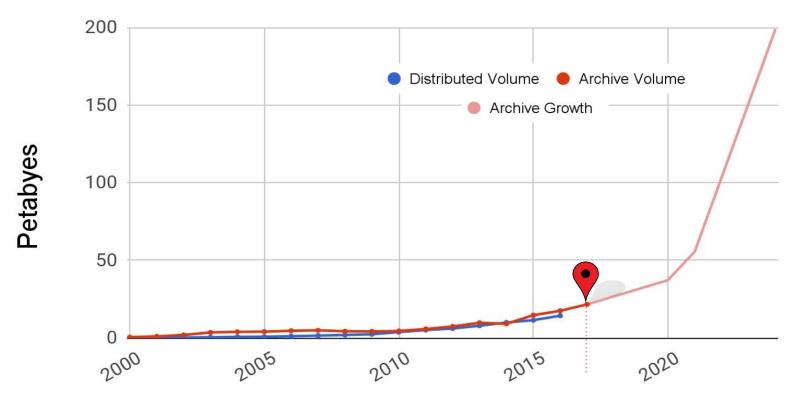Ian Schuler, Alireza Jazayeri - Development Seed

EARTHDATA
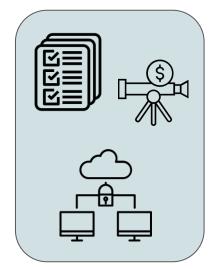**EOSDIS** NASA'S EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM

EOSDIS in Context

Distribute → Research, Applications, Education

Transform*

Archive

Process

Capture and Clean

*Subset, reformat, reproject

ASF DAAC
SAR Products, Sea Ice, Polar Processes

U. of Wisc.
SNPP Atmosphere

SEDAC
Human Interactions in Global Change

LPDAAC
Land Processes and Features

CDDIS
Crustal Dynamics Solid Earth

GSFC
SNPP, MODIS, OMI, OBPG

NCAR, U. of Co.
MOPITT

JPL
MLS, TES, SNPP Sounder

OB.DAAC
Ocean Biology and Biogeochemistry

PO.DAAC
Ocean Circulation Air-Sea Interactions

NSIDC DAAC
Cryosphere, Polar Processes

LAADS/MODAPS
Atmosphere

GHRC
Hydrological Cycle and Severe Weather

GHRC
AMSR-U, LIS

ASDC
Radiation Budget, Clouds, Aerosols, Tropo Composition

SIPS

DAAC

3

EOSDIS Holdings and Projected Growth

Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas

Compliance, Security, Cost Tracking

Core Archive Functionality and Processing
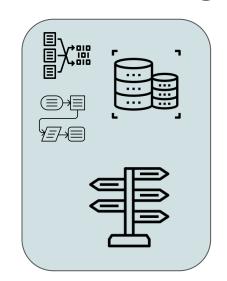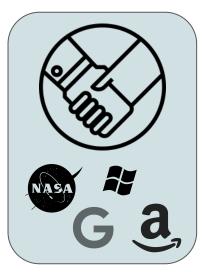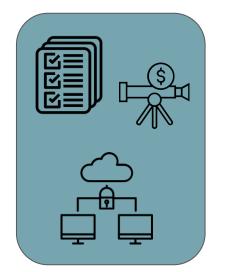
End-User Application Migration

Pursuing Cloud Partnerships

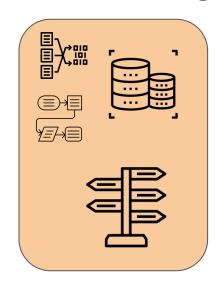# Past 24 Months: Focused on evaluation and planning for a cloud migration in 4 areas
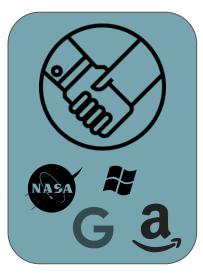

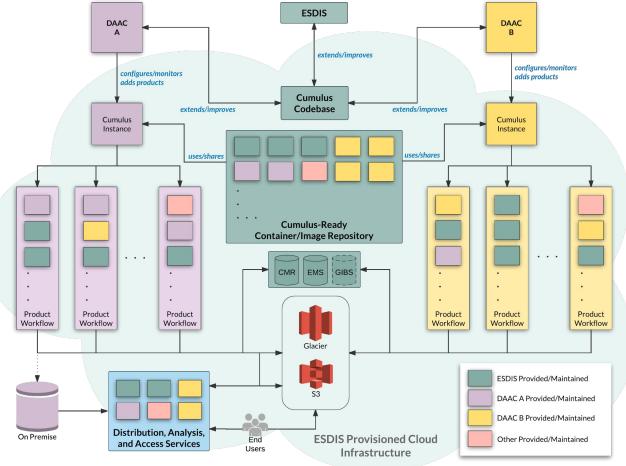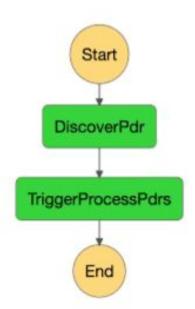Compliance, Security, Cost Tracking

Core Archive Functionality and Processing

End-User Application Migration
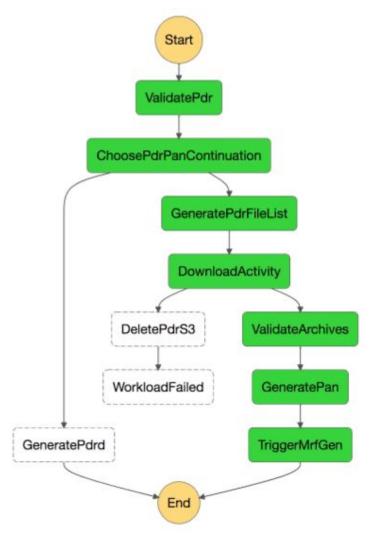
Pursuing Cloud Partnerships
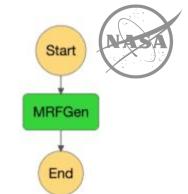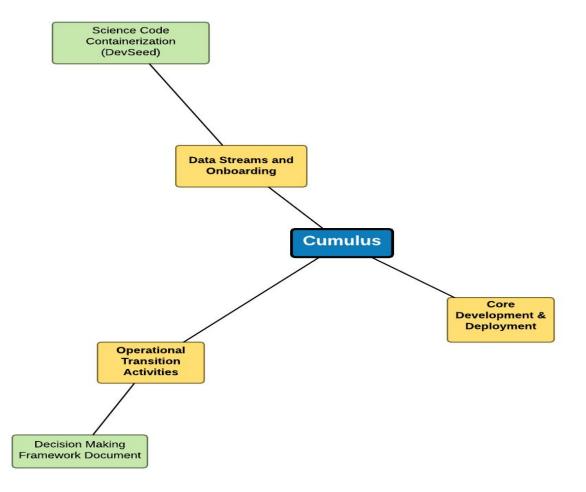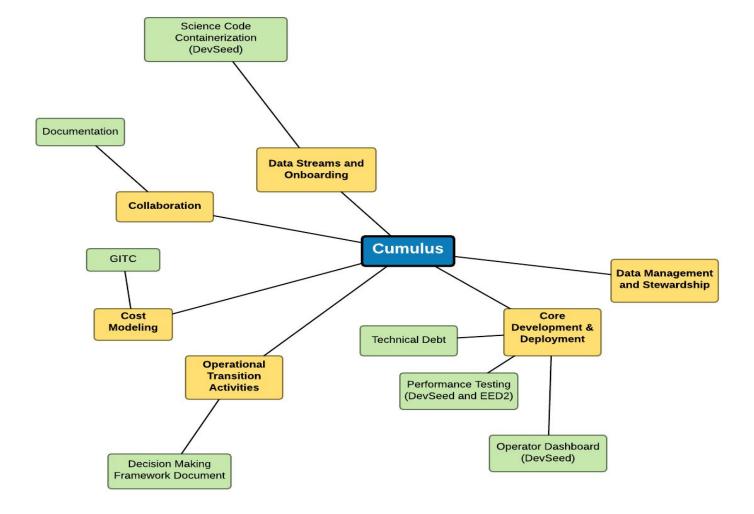
Example Step Function for GIBS in the Cloud
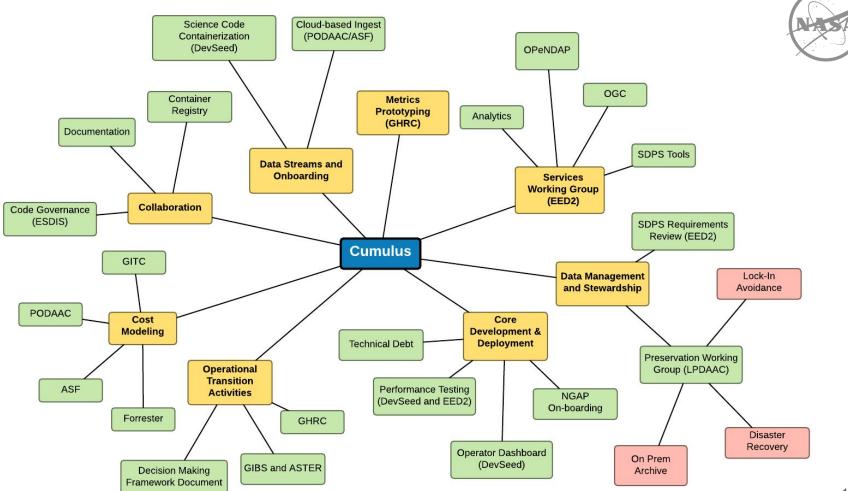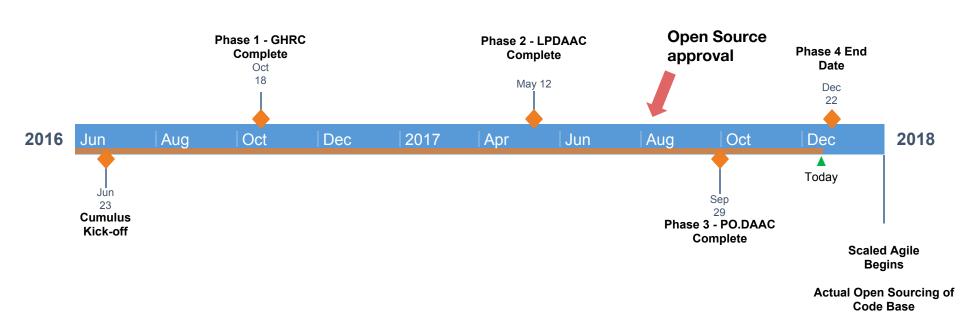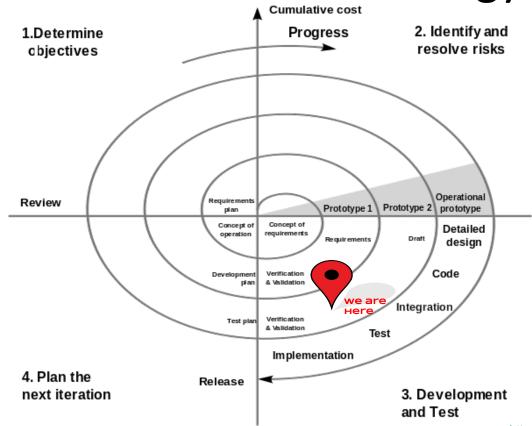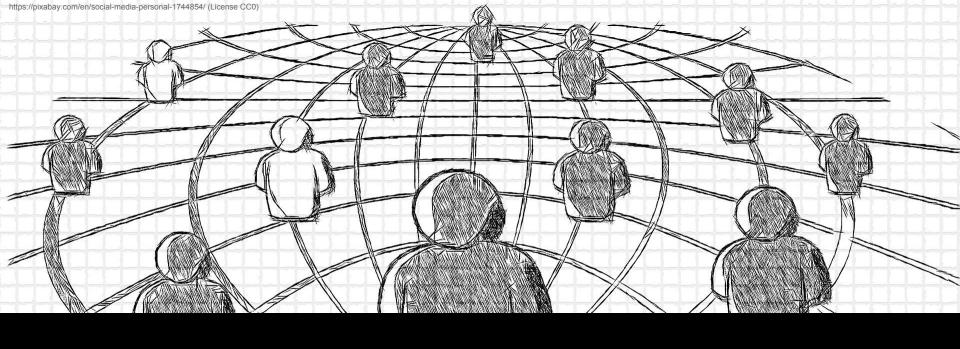
Looking at the Past 18 Months

Phase 1 - GHRC Complete
Oct 18

Phase 2 - LPDAAC Complete
May 12

Open Source approval

Phase 4 End Date
Dec 22

2016

| Jun | Aug | Oct | Dec | 2017 | Apr | Jun | Aug | Oct | Dec |

2018

Jun 23
Cumulus Kick-off

Today

Sep 29
Phase 3 - PO.DAAC Complete

Scaled Agile Begins

Actual Open Sourcing of Code Base
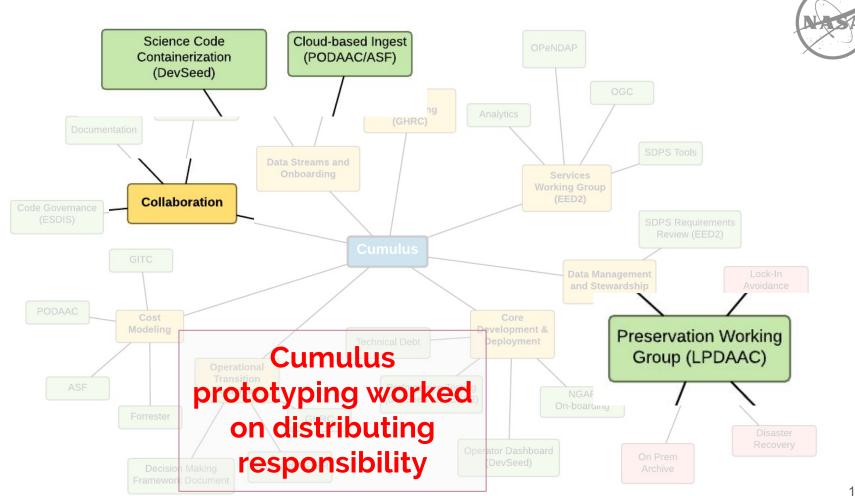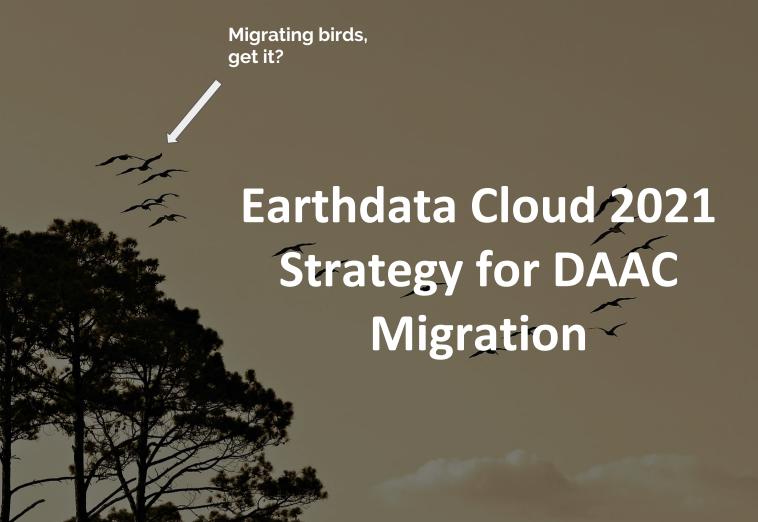
Cumulus Phased Methodology

1.Determine objectives

Cumulative cost

Progress

2. Identify and resolve risks

Review

Requirements plan

Prototype 1 | Prototype 2 | Operational prototype

Concept of operation | Concept of requirements | Detailed design

Requirements | Draft

Development plan | Verification & Validation | Code

we are Here

Integration

Test plan | Verification & Validation

Test

Implementation

Release

4. Plan the next iteration

3. Development and Test

https://en.wikipedia.org/wiki/Spiral_model

So, what does this mean in practical terms for EOSDIS?

Science Code Containerization (DevSeed) — Cloud-based Ingest (PODAAC/ASF) — OPeNDAP — OGC — Analytics — Documentation — Data Streams and Onboarding — (GHRC) — Services Working Group (EED2) — SDPS Tools — Collaboration — Code Governance (ESDIS) — Cumulus — SDPS Requirements Review (EED2) — GITC — Data Management and Stewardship — Lock-In Avoidance — PODAAC — Cost Modeling — Technical Debt — Core Development & Deployment — Preservation Working Group (LPDAAC) — ASF — Operational Transition — NGAP On-boarding — Forrester — GHRC — Disaster Recovery — Decision Making Framework Document — Operator Dashboard (DevSeed) — On Prem Archive

Cumulus prototyping worked on distributing responsibility

Earthdata Cloud 2021 Strategy for DAAC Migration

Migrating birds, get it?

16

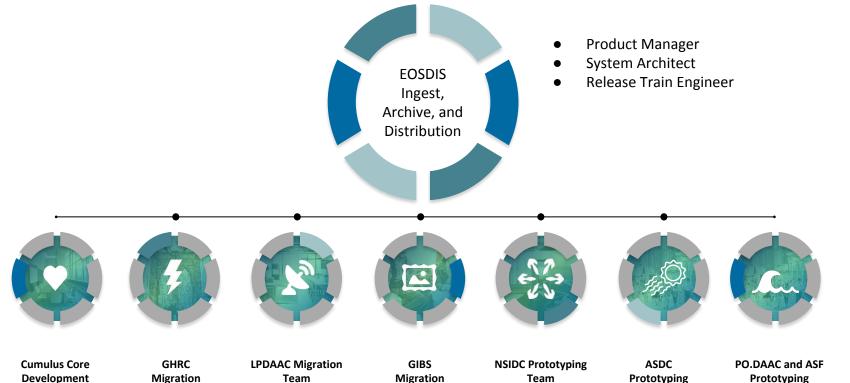# Using SAFe®* to align Earthdata Cloud 2021 Migrations

*Yes, I am aware of the huge number of buzzwords on image above. And it is less "agile".
But this project involves 100s of people and this system has proven extremely useful.
http://www.scaledagileframework.com/program-increment/

DAAC Cloud Migration Organization January

EOSDIS Ingest, Archive, and Distribution

- Product Manager
- System Architect
- Release Train Engineer

Cumulus Core Development

GHRC Migration Team

LPDAAC Migration Team

GIBS Migration Team

NSIDC Prototyping Team

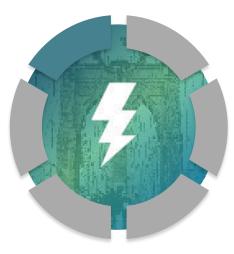**ASDC Prototyping Team**

PO.DAAC and ASF Prototyping Teams
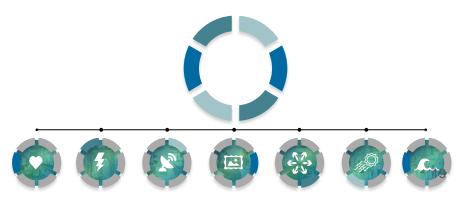
DAAC Team Organization

DAAC-owned Teams

- Led from internal DAAC resources
- Development, operators, testing
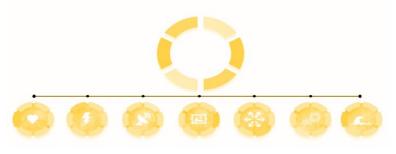- Requests external support from Core team for specific tasks

Cumulus Platform Core

- Mainly ESDIS contracted development/design resources
- Working on smoothing out the delivery, versioning, on-boarding
- Aiding DAAC teams at specific intervals (development help, face to face meetings, etc) "We need help integrating xyz, etc"
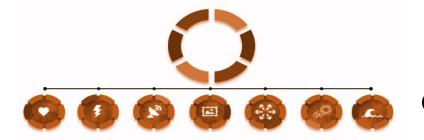
Ingest, Archive, and Distribution is only one of 3 three potential scaled agile teams

Data Usage (Clients and Services)

Ingest, Archive, Distribution

Cloud Hosting Platform

DAAC Cloud Migration PI Roadmap

OCT - DEC

- Planning Activities
- Process alignment
- NSIDC Initial Engagement
- GIBS Performance Testing

APR - JUN

- Operational Data Onboarding Begins (LPDAAC, GHRC, GIBS)
- Service Integration Begins (LPDAAC)*
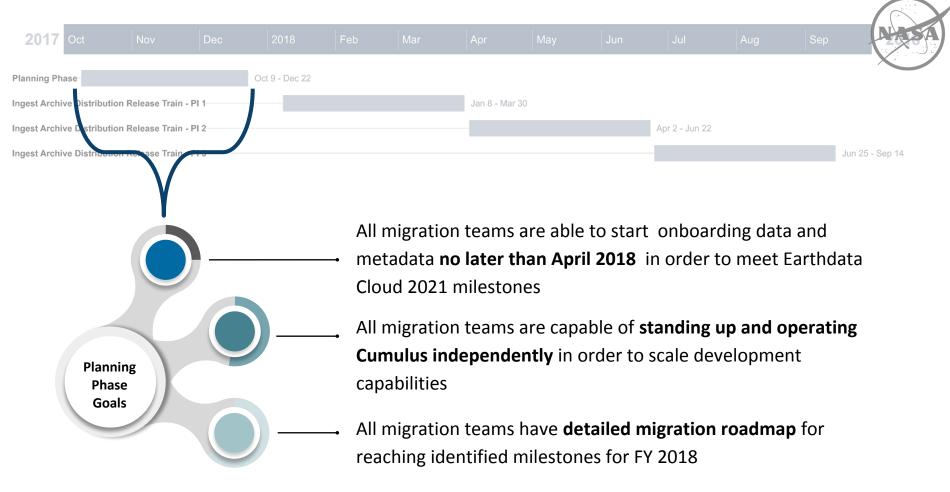- ASDC Initial Engagement

JAN - MAR

- System Evolution
- Collaboration Coordination
- NSIDC New Data Streams

JUL - SEP

- Operational System Testing (LPDAAC, GHRC, GIBS)
- Continued Service Integration
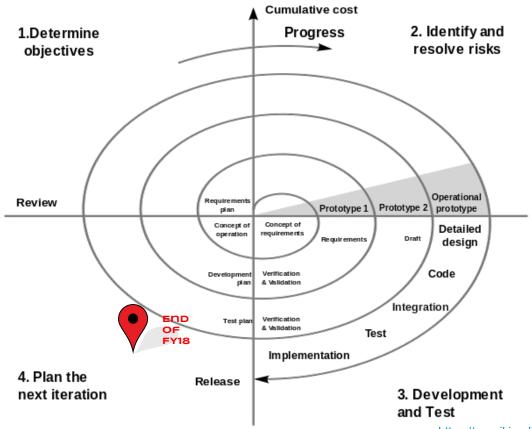- ASDC New Data Streams

Q1 and Q2 2018 Migration Goals

- All migration teams are **capable of contributing and testing code/tasks/services to Cumulus** core in order to scale development capabilities **(March 2018)**

- **Cumulus Core Evolution and Cross-DAAC System Engineering Activities**
  - Updated, cloud-native **requirements** documentation
  - Cloud-native **metrics** system
  - **Backup and recovery** procedures for archived data, workflows, and tasks
  - **Distribution API and Egress Shaping**

Ingest/Archive/Distribution in FY18

https://en.wikipedia.org/wiki/Spiral_model

Overview

Code Governance Document: http://bit.ly/2h88Nm2

DAAC/GIBS Data stream status: http://bit.ly/2j1JZNq

Existing Cumulus Tasks (reusable components): http://bit.ly/2yqhGyJ

Cumulus Docs (very much in work) https://cumulus-nasa.github.io/

Cumulus Code Base: https://github.com/cumulus-nasa

What about Code Governance?

Because this as an **organizational** shift,
not just a technology shift

Drafting a contribution guide

1. Provide working definitions of the high-level components of the Cumulus system, including specifying which of those components are governed by this document.
2. Establish roles and responsibilities for contributions to Cumulus NASA's EOSDIS
3. Identify key communication flows, as well as information on documentation, testing and deployment paradigms
4. Outline high-level process expectations for Cumulus contributions and provide example process flows for these contributions

Project Structure

Project Management Board

sets the vision for →

Project Leadership Team

provides the "what" for →

Technical Leadership Team

guides the "how" for →

Committers

Contributors

This diagram lays out the relationships between the various roles in the team, showing how each team interacts at a high-level.

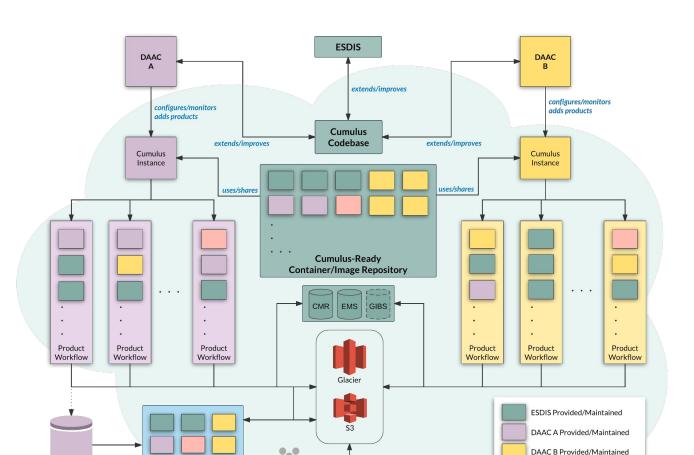As we scale (up or down) we can adjust/combine/tailor these roles. We can adapt to other projects/systems.

28
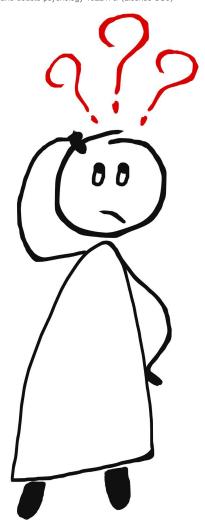
Questions?

katie.baynes@nasa.gov

DAAC	Distributed Active Archive Center
SIPS	Science Investigator-led Processing System
EOSDIS	Earth Observation System Data and Information System
CMR	Common Metadata Repository
EMS	EOSDIS Metrics System
NGAP	NASA Compliant General Application Platform
ESDIS	Earth Science Data and Information System
GIBS	Global Imagery Browse Services
GHRC	Global Hydrology Resource Center
LP DAAC	Land Processes DAAC
PO DAAC	Physical Oceanography DAAC
ASDC	Atmospheric Science Data Center
NSIDC	National Snow and Ice Data Center
SAFe®	Scaled Agile Framework®
ASF	Alaska Satellite Facility

Acronym List!